Nucleic Acids Research, 1997, Vol. 25, No. 13 **2661–2671**

Identification and characterization of the human *XIST* gene promoter: implications for models of X chromosome inactivation

Brian D. Hendrich+, Robert M. Plenge and Huntington F. Willard*

Department of Genetics and Center for Human Genetics, Case Western Reserve University School of Medicine, Cleveland, OH 44106-4955, USA

Received February 13, 1997; Revised and Accepted May 15, 1997

DDBJ/EMBL/GenBank accession nos U50908–U50912

ABSTRACT

The *XIST* gene in both humans and mice is expressed exclusively from the inactive X chromosome and is required for X chromosome inactivation to occur early in development. In order to understand transcriptional regulation of the *XIST* gene, we have identified and characterized the human *XIST* promoter and two repeated DNA elements that modulate promoter activity. As determined by reporter gene constructs, the *XIST* minimal promoter is constitutively active at high levels in human male and female cell lines and in transgenic mice. We demonstrate that this promoter activity is dependent *in vitro* upon binding of the common transcription factors SP1, YY1 and TBP. We further identify two *cis*-acting repeated DNA sequences that influence reporter gene activity. First, DNA fragments containing a set of highly conserved repeats located within the 5′-end of *XIST* stimulate reporter activity 3-fold in transiently transfected cell lines. Second, a 450 bp alternating purine–pyrimidine repeat located 25 kb upstream of the *XIST* promoter partially suppresses promoter activity by ~70% in transient transfection assays. These results indicate that the *XIST* promoter is constitutively active and that critical steps in the X inactivation process must involve silencing of *XIST* on the active X chromosome by factors that interact with and/or recognize sequences located outside the minimal promoter.

INTRODUCTION

X chromosome inactivation results in random transcriptional inactivation of one of the two X chromosomes present in normal, female mammalian cells. This process allows mammals to achieve dosage equivalence of most X-linked genes between females, who normally have two X chromosomes, and males, who normally have one X chromosome. The major genetic locus proposed to control the X chromosome inactivation process is the X inactivation center (*XIC*). *XIC* is defined as a region of the X chromosome from which a currently ill-defined inactivation signal exerts its effect *in cis* along the chromosome; derivative X chromosomes lacking this *XIC* are unable to become inactivated (1–5). Human *XIC* has been localized to a <1 Mb region within band Xq13.2, while murine *Xic* maps to the homologous location on the murine X chromosome (6–8). A second genetic locus known to influence the X inactivation process in mice is the X chromosome controlling element (*Xce*; 9). Different alleles at the *Xce* locus influence the degree of randomness of the X inactivation process (10,11). Localization of the *Xce* locus to within the *Xic* region in mice has led to the speculation that *Xce* and *Xic* are synonymous loci (12,13).

The *XIST* gene, whose product is a non-coding nuclear RNA, has been implicated strongly in the process of X chromosome inactivation due to its map location within *XIC* and its unique inactive X-specific transcription pattern (14). Expression of the *XIST* gene is tightly correlated with the presence of an inactive X chromosome and *XIST* transcripts are found closely associated with the inactive X chromosome in interphase nuclei (15,16). Transcripts from the *Xist* gene in mouse are found at high levels a full day before X inactivation is believed to occur in murine development, an observation that is consistent with *XIST*/*Xist* having an initiating role in the X inactivation process (17). Recently, a targeted deletion of the *Xist* gene was created in murine embryonic stem (ES) cells. The X chromosome carrying the mutant *Xist* allele was unable to be inactivated, providing direct evidence that expression of the *Xist* gene is necessary for X inactivation to occur in ES cells as well as in chimeric mouse embryos (18). Notwithstanding their apparent inability to inactivate the X chromosome carrying the mutation, cells carrying the targeted *Xist* allele appeared to carry out early steps in the inactivation process normally, i.e. both recognition of the number of X chromosomes present and random choice of which X was to become inactive were unaffected by the *Xist* mutation. Thus the deletion, which comprised part of the *Xist* promoter and part of the first exon, does not affect these steps. These results imply a spatial separation between sequences responsible for different steps in the initiation of X inactivation (19).

Transgenic mice have been created in several laboratories in which yeast artificial chromosomes (YACs) containing portions of the *Xic* region were integrated into ectopic sites in the murine genome (20–23). While it was possible to achieve expression of

*To whom correspondence should be addressed. Tel: +1 216 368 1617; Fax: +1 216 368 3030; Email: hfw@po.cwru.edu

+Present address: Institute of Cell and Molecular Biology, The University of Edinburgh, Darwin Building, King's Buildings, Edinburgh EH9 3JR, UK

the *Xist* gene in some instances, this did not always result in detectable transcriptional silencing of the host chromosome. In fact, evidence for *Xist* expression and inactivation of the host autosome has been presented for only a single multicopy transgene (22,23). In contrast, transgenic cell lines containing two to eight copies of a *Xist* cosmid integrated into autosomes were able to both express *Xist* from the transgene and repress transcription of a reporter gene *in cis* (24). Similarly, both *XIST*/*Xist* expression and spread of inactivation are readily observed when intact *XIC*/*Xic* is involved in a translocation with autosomal material (25). One possible interpretation of these seemingly contradictory results is that the *Xist* gene is subject to complex regulatory mechanisms requiring sequences that either are not present or are not maintained in a proper context in the transgenes in some studies.

Combined, the available data indicate that *XIST*/*Xist* expression and accumulation of *XIST*/*Xist* RNA are involved in the initiation of X chromosome inactivation. Further, different levels of steady-state *Xist* RNA have been reported in mice and in differentiated ES cells carrying different alleles at the *Xce* locus (26,27), suggesting a possible link between the *Xce* locus and the *Xist* transcriptional regulatory elements. Thus, characterization of the *XIST* promoter should provide insights into the nature of *XIST* transcriptional regulation, initiation of X inactivation, the nature of any interaction between the promoter and other sequences within *Xce* and/or *XIC* and the identities of other factors involved in the X inactivation process.

In order to understand the transcriptional regulation of *XIST*, it is necessary to first determine what factors are required for its transcription and then to identify other elements that influence the ability of the transcriptional machinery to identify the promoter and initiate transcription. Towards this end, in this paper we describe identification and characterization of the *XIST* minimal promoter. We have identified binding sites for common transcription factors within the minimal promoter sequence and, in addition, describe two *cis*-acting sequences that modulate minimal promoter activity.

MATERIALS AND METHODS

Clone isolation

Screening of human λ phage libraries, PCR, cloning and isolation of primate DNAs were carried out as described (15,28). A murine genomic λ phage clone was obtained from a genomic library constructed from a YAC containing the murine *Xist* locus (YAC 4B-2, a gift of Dr Phil Avner). DNA was prepared from a yeast culture containing the *Xist* YAC and partially digested with *Mbo*I to provide DNA in the range 10–20 kb. DNA was ligated into predigested, phosphatased λ DASH II vector arms (no. 246211; Stratagene) and packaged using Gigapack II Gold packaging extracts (no. 247612; Stratagene) according to the manufacturer's instructions. Approximately 10^6 phage were screened with the murine cDNA probe (28) at a final wash stringency of 0.1% SDS, 0.1× SSC at 65°C. Hybridizations were carried out at 65°C in a hybridization solution of 10% dextran sulfate, 1 M NaCl, 1% SDS.

The lepine cDNA clone was isolated from a female rabbit liver (no. TL 1006a; Clontech) cDNA library generated by oligo(dT) priming. One lepine cDNA clone was obtained by screening ~10^7 primary plaques with the 5′-most human *XIST* cDNA probe (Hbc1a) at a final wash stringency of 0.5% SDS, 50 mM Tris–HCl, pH 8.6, and 0.5 M NaCl at 65°C. An overlapping lepine λ genomic clone was obtained by screening a rabbit genomic library (no. TL1008j; Clontech) at a final wash stringency of 0.1% SDS and 0.1× SSC at 65°C with the lepine cDNA clone. An equine genomic λ clone was isolated from a male horse library (no. 946701; Stratagene) with the Hbc1a probe as described above. Sequences homologous to the human *XIST* 5′-region were subcloned and sequenced.

Contig generation, sequence determination and analysis

Nucleotide sequence of cDNAs was determined on double-stranded templates using vector- and gene-specific primers as described (15,28), in most instances using an Applied Biosystems fluorescence sequencer (ABI model 373A or 377, with V1.1.1 or V1.2 sequence analysis software). Contig assembly and sequence analysis was performed using either the GeneWorks (Intelligenetics) or LaserGene (DNAStar) DNA analysis software. Sequence comparisons were performed using the GeneWorks DNA alignment program. Database searches were performed using the BLAST network service at the National Center for Biotechnology Information (NCBI). The GRAIL 2 and XGRAIL v1.2 programs were used to evaluate protein coding potential (29).

Cell culture, transfection and luciferase assays

The female embryonic kidney cell line (293) and male fibrosarcoma cell line (HT 1080) were purchased from ATCC (CRL-1573 and CRL-7951 respectively). Promoter elements were cloned into GeneLight vectors (Promega). Approximately 10^5 cells were transiently transfected using 1 μg plasmid and 1.5 μg lipofectin in a total volume of 250 μl serum-free medium overnight. Medium was changed after ~16 h and cells were harvested after 48 h using Cell Lysis Reagent (Promega) according to the manufacturer's instructions. Aliquots of 20 μl cell lysate were used to measure luciferase activity by addition of 100 μl Luciferase Assay Reagent (Promega), followed by luminescence quantitation in a TD-20e luminometer (Turner Designs).

All transfections were carried out in duplicate. Each luciferase reading was normalized to the average minimal promoter activity for each experiment. Two tailed *t*-tests and other statistical computations were done using software supplied with Microsoft Excel v. 5.0.

Site-directed mutagenesis

PCR conditions and selection of PCR primers was as described (28). Each mutagenesis construct was created by making two primers facing in opposite directions whose 3′-ends lie six bases apart. An *Eco*RI linker was added to the 3′-ends of each primer. PCR was carried out with the mutagenesis primer in conjunction with G7R (GAAGTTGTGACTCCTGGTCT) for the 5′-facing primers or G10R (GAGAGATCTTCAGTCAGGAAG) for the 3′-facing primers. G7R contains an *Xba*I site, while G10R contains a *Bgl*II site. The two PCR products were co-precipitated and then resuspended in Universal restriction enzyme buffer (Stratagene). The reactions were then digested with *Eco*RI, *Xba*I and *Bgl*II simultaneously. Digestion products were then co-precipitated with plasmid pGLB which had been digested with *Nhe*I and *Bgl*II. Digestion products were then ligated together at 16°C overnight. The ligation reaction was transformed into DH5α (BRL) or One-Shot competent cells (Invitrogen) in the presence

of ampicillin. Ampicillin-resistant colonies were picked into 150 µl TB/Amp in a 96-well plate and grown overnight. Samples of 2 µl of culture were PCR amplified with vector primers GLP1 (TGTATCTTATGGTACTGTAACTG) and GLP2 (CTTTATGTTTTTGGCGTCTTCCA), digested with *Eco*RI and electrophoresed on 2% agarose gels. Positive colonies were further analyzed by sequencing.

The pGLXB construct was subjected to *Bal*31/exonuclease III directional deletions according to the manufacturer's instructions (Stratagene) to generate the –129 and –72 constructs.

Electrophoretic mobility shift assays

PCR products or double-stranded oligonucleotides (SP1, TFIID, YY1 and SP1 mutant oligonucleotides; Santa Cruz Biosystems) were end-labeled with [γ-^{32}P]dATP using polynucleotide kinase (New England Biolabs) and purified using Nuc-Trap columns (Stratagene) according to the manufacturer's instructions. Approximately 10^4 c.p.m. (100 pg) labeled oligonucleotide were incubated with 5 µg HeLa nuclear extract (Promega) or recombinant SP1 protein (Promega) in a final concentration of 10 mM Tris, pH 8, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 2 mM DTT, 50 µg/ml BSA, 2 µg/ml sonicated herring sperm DNA, 100 mM KCl, 10% glycerol and 0.3 µg/ml poly(dI·dC) at room temperature for 30–60 min. Binding reactions were loaded onto 6% non-denaturing polyacrylamide gels electrophoresed in 0.5× TBE at room temperature. Gels were transferred to Whatman paper, dried and exposed to X-ray film. Bandshifts with probe 2 (L/S12 region) were carried out as described above in a binding buffer consisting of 20 mM Tris, pH 8, 2 mM DTT, 80 mM KCl, 10 mM $MgCl_2$, 10% glycerol. Non-denaturing acrylamide gels and electrophoresis buffer contained 0.02% NP-40 and 4 mM $MgCl_2$. These gels were electrophoresed at room temperature or 4°C. Antibodies (polyclonal anti-Sp1, no. sc-59; polyclonal anti-TBP, no. sc-204; monoclonal anti-TBP, no. sc-421; polyclonal anti-YY1, no. sc-281; polyclonal anti-USF, no. sc-229; Santa Cruz Biotechnology) were added to the binding reactions in supershift experiments and incubated at room temperature for at least 30 min before electrophoresis.

Transgenic mice

Transgenic mice were created at the Transgenic Mouse Facility in the Department of Genetics at Case Western Reserve University by pronuclear injection. Transgenic embryos were harvested at E9.5–E13.5 in ice-cold PBS and a portion of each embryo was then removed for genotype analysis by PCR (30) using transgene-specific primers [transgenes XH and HH, primers G10 (CTTCCTGACTGAAGATCTCTC) and GLP2 (see above); transgene G6H6, primers G6 (TACTCTTCCACTCACTTTTC) and H6 (AGAGAGTGCAACAACCCACA)] and primers for the *Sry* gene to determine the gender of each embryo (Sry1, GATCAGCAAGCAGCTGGGAT; Sry2, TTTGGGTATTTCTCTCTGTG). The remainder was homogenized in Cell Lysis Reagent (Promega). The homogenate was centrifuged for 2 min at maximum speed in a microcentrifuge and 20 or 60 µl supernatant were assayed for luciferase activity as described above. Total protein concentration of supernatant was determined using a Bradford assay according to the manufacturer's instructions (BioRad).

Database accession numbers

Sequences described here have been deposited into the GenBank sequence repository. The accession numbers for the human, mouse, rabbit and horse *XIST*/*Xist* sequences are U50908–U50911 respectively and that for the sequence containing the PuPy repeat is U50912.

RESULTS

The *XIST* minimal promoter

Conservation of promoter sequences. To identify potentially important sequences within the *XIST* 5′-region, sequences upstream of the murine, lepine and equine *Xist* genes were cloned, sequenced and compared with the corresponding human sequence. In total, 850 bp of lepine sequence, 2010 bp from the murine locus and 3838 bp of equine upstream sequences were compared with 6475 bp of human sequence 5′ of the *XIST* gene itself. No significant sequence homology was detectable more than ~100 bp upstream of the transcription start site (data not shown). No other genes or pseudogenes were identified in any of the upstream sequences using standard database searching and gene finding programs (29,31) or, in the case of the human sequence, using RT-PCR expression analysis (data not shown).

Within the segments that showed clear evidence of homology, comparison of the immediate upstream portions of the four sequences is shown in Figure 1. A region of elevated conservation among the four sequences is present within ~100 bp of the transcription initiation site. The region of elevated conservation in this upstream sequence (–101 to –1) is 74, 78 and 81% identical between human and mouse, human and rabbit and human and horse respectively and is comparable with levels of identity observed in the 5′-end of the RNA itself (bases +1 to +308) for the same three comparisons (Fig. 1; 32; B.D.Hendrich, PhD thesis, Stanford University). This high degree of conservation in the immediate upstream region suggests that this sequence is important in *XIST* function and suggests this region as a candidate for *XIST* promoter sequences. The sequence alignment reveals no conserved TATA or CCAAT sequences. The human sequence does, however, contain a consensus SP1 binding sequence (33) located from position –49 to –54, which is completely conserved among ape and Old World monkey *XIST* genes (data not shown) and shows partial conservation in the murine and lepine sequences (Fig. 1). In addition, the conserved sequences around the transcription start site resemble the consensus binding site for the initiator protein YY1 (34–36).

Identification of minimal promoter activity. To identify the *XIST* promoter functionally, a number of constructs were prepared in which differing amounts of DNA including the transcription start site were cloned just upstream of a promoterless luciferase gene in the pGLB vector (Fig. 2). Promoter activity was assayed by transiently transfecting plasmid constructs into both male (HT 1080, fibrosarcoma) and female (293, embryonic kidney) human cell lines and detecting luciferase activity in cell lysates (see Materials and Methods). Constructs containing ~100 bp to ~1 kb of upstream sequence each conferred high levels of luciferase activity in transient transfections (Fig. 2). Approximately equivalent levels of activity were detected in male and female cells; all subsequent experiments were performed with female 293 cells only, for convenience.

```
Human Promoter    TGGTATGTCT TTGCTTTCTT GAAATTTCCC TTTTTTGTCC TTACTGGGTA AATTTTGAAC CAACCAAATC ACAAAGATGT   -95
Mouse Promoter    TACGTACCTC CATCTTTAT- TCA-TTTTAA TTTTTTTATA ATA-TAGTTA GACCT--AAA GGTCC-AAT- ---AAGATGT   -93
Rabbit Promoter   ATCCCCCTGT TTTCTTCAGC TCT-CCTCAT TTTGCTGGT- -TACTGGGTA CAGCTTGAGT CGACCAAACC TCTAAGATGT   -93
Horse Promoter    CTAATATGTT GCTGTTTTCT TCAAATACCC CTTTTTCTTG TTCCTAAGTA AATTTGGGGC CGGCTAAACC ACAAAGATAT   -99

                                  L/S7  L/S6  L/S5                                          L/S12
Human Promoter    CCGGCTTTCA ATCTTC-TAG GCCACGCCTC TTATGCTCTC TCCGCCCTCA GCCCCCCC-- --TTCAGTTC TTAAAGCGCT   -20
Mouse Promoter    CAGAATTGCA ATCTTT-GTG GCCACTCCTC TTCTGGTCTC TCCGCCTTCA GCGCCGCG-- -GATCAG--- TTAAAGGCGT   -20
Rabbit Promoter   CGGGCCTTCA ATCAGCGGGG GCCACGCCTC TTGT----TC ACCCCGCCCC CAACCCCC-- -CCTCAATTC TTAAAGCGCT   -20
Horse Promoter    ATGGCTTTCA ATCTTC-TAG GCCACGCCTC TTCTACTTCC TCCACCCCCC AGTCCCCTCC CCCTCACTTC TTAAAGCGCT   -20

                                  L/S17
Human Promoter    GCAATTCGCT GCTGCAGCCA TATTTCTTAC TCTCTCGGGG CTGGAAGCTT CCTGACTGAA GAT-CTCTCT GCACTTGGGG   +60
Mouse Promoter    GCAACGGCTT GCTCCAGCCA TGTTTGCTCG TTTCCCGTGG ATGTGCGGTT CTT--CCGTG GTTTCTCTCC AT-C-TAAGG   +57
Rabbit Promoter   GCATTTTGCT GCAGCAGCCA TATTTCTACT TCTCCCGAGG TTGGAAGCTC GCTAGCCATA GTT-CTTTCT GTACTTATGG   +60
Horse Promoter    GCAATTTGCT GCTGCCGCCA TATTTTTTCT TTTCCT-AGG GTGGAAACTT GCTAACACTT GAT-CTCTTT GCCCGTGTGG   +59
                                  *
```

Figure 1. Alignment of human, murine, lepine and equine *XIST*/*Xist* sequences. *XIST*/*Xist* 5′-end and upstream sequences. The sequence of the *XIST*/*Xist* promoter region is given from human, mouse, rabbit and horse. Bases conserved in all four sequences are shaded. The location of the transcription initiation site is indicated with an asterisk. Human sequence positions with respect to the transcription start site (+1) are shown at right and above the human sequence. The consensus SP1 binding site (–49 to –54) and the near-consensus YY1 binding site (–3 to +6) are indicated in bold. Bases which when mutated in mutagenesis constructs show a significant decrease in promoter activity are indicated above the sequence with the names of the respective constructs.

To further define the minimal promoter, deletions were made in the pGLXB plasmid to provide constructs containing 129, 93 and 72 bases upstream of the transcription start site respectively (Fig. 3A). While the –129 and –93 constructs generated luciferase levels approximately equal to those generated by the full promoter construct (pGLXB), the –72 plasmid gave only background levels of luciferase activity. Considering all of the data, the minimal promoter is thus defined to lie between positions –93 and +31. Genomic fragments containing the homologous sequences from murine, lepine and equine DNA (Fig. 1) also showed promoter activity in this assay (Fig. 2).

To investigate the activity of the *XIST* promoter *in vivo*, we created transgenic mice in which the human *XIST* promoter was used to drive a luciferase reporter gene. Three different transgenes were created in which different parts of the *XIST* promoter were used to drive the reporter gene [constructs GLHH and GLXH (Fig. 2B) and G7RH6 (Fig. 5B)]. Multiple lines were analyzed for each transgene and all lines were analyzed through multiple generations for potential imprinting or sex-specific effects. No such effects were seen for any of the transgenic lines and all transgenes segregated in an autosomal pattern. In total, 11 different lines were obtained, nine of which exhibited detectable levels of luciferase activity. Although expression levels varied up to several thousand-fold (even when transgene copy number was taken into account), this is consistent with the *XIST* minimal promoter being constitutively active and is in contrast to some of the results obtained when *Xist*-containing YACs were used for transgenesis (20–22; see Discussion).

Saturation site-directed mutagenesis. In order to identify potentially important sequences within the *XIST* minimal promoter, the entire minimal promoter was subjected to saturation site-directed mutagenesis. Every six bases from position –108 to +11 were sequentially replaced by an unrelated hexamer, cloned in front of the promoterless luciferase gene and transiently transfected into tissue culture cells (see Materials and Methods; Fig. 3). While the majority of mutagenesis constructs retained promoter activity, five clones failed to produce high levels of luciferase activity when compared with the unmutated promoter. Two mutant constructs, L/S6 and L/S7, define a 12 bp region extending from –78 to –67 (Fig. 1). Construct L/S5 also showed decreased promoter activity, though the effect was much less than that seen for L/S6 and L/S7. The L/S12 construct also showed reduced activity and disrupts the highly conserved sequence TTAAAG. Lastly, the sequence mutated in L/S17 contains the transcription start site at positions –1 to +5 and overlaps the consensus YY1 binding sequence (Fig. 1).

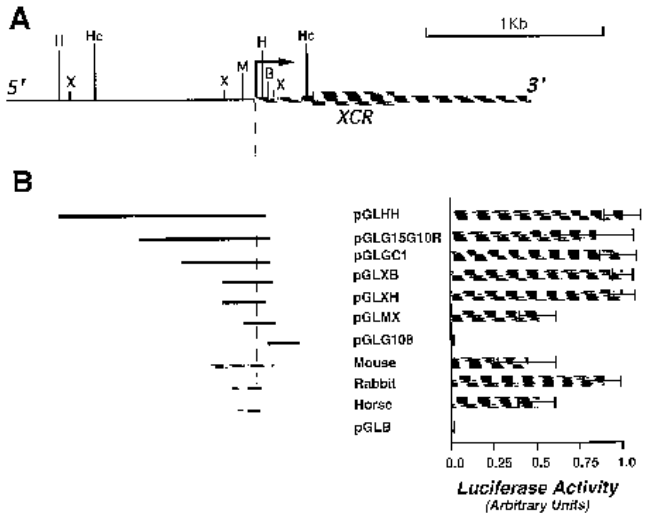

Figure 2. Identification of the *XIST* minimal promoter. (**A**) Partial restriction map of the 5′-end of the human *XIST* gene, with the gene indicated as a thick line. The transcription initiation site is shown as a large arrow, with transcription proceeding to the right. The location of the XCR sequences is indicated by the gray box. Restriction enzyme sites are as follows: B, *Bam*HI; H, *Hin*dIII; Hc, *Hin*cII; M, *Msp*I; X, *Xba*I. (**B**) Different portions of the 5′-end of the human *XIST* gene were cloned into the multiple cloning site (MCS) of a luciferase reporter construct and transiently transfected into female embryonic kidney (293) cells. The amount of 5′ material cloned into each construct is indicated as a black line. The location of the transcription initiation site in each construct is indicated with a dotted line. Luciferase levels are plotted in arbitrary amounts, normalized to that produced by the pGLXH construct. Homologous sequences in mouse, rabbit and horse (gray lines) were also shown to promote transcription of the reporter gene. The nucleotide positions contained within each *XIST* construct are: pGLHH, –1187 to +31; pGLG15G10R, –816 to +46; pGLGC1, –537 to +46; pGLXB, –211 to +46; pGLXH, –211 to +31; pGLMX, –93 to +71; pGLG108, +29 to +497.

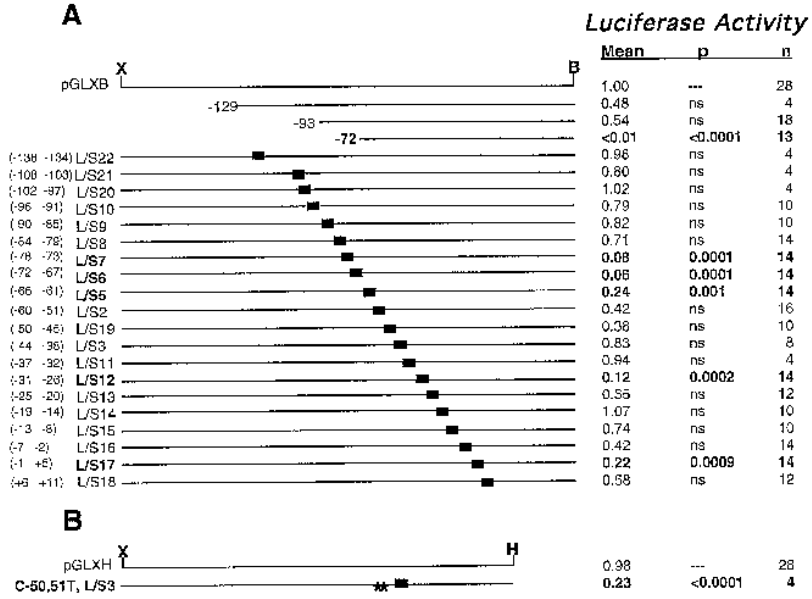

Figure 3. Site-directed mutagenesis of the human *XIST* minimal promoter. Constructs were transiently transfected into 293 cells and luciferase activity measured as in Figure 2. Mean luciferase values are expressed as fractions of wild-type minimal promoter activity [pGLXB in (A), pGLXH in (B)], with the number of experiments contributing to the mean shown in parentheses. Statistical significance (*P*) was determined using a two tailed *t*-test. Constructs showing a significant decrease in promoter activity ($P \leq 0.001$) are indicated in bold. (**A**) Deletion constructs are shown as partial lines with numbers corresponding to the amount of upstream promoter sequence. Linker/scanner (L/S) constructs are indicated below, with the box corresponding to the substitution of six bases by an*Eco*RI site (GAATTC). The positions of the substituted bases with respect to the transcription initiation site are indicated to the left of each construct. Construct L/S14 contains a*Bam*HI site (GGATCC) rather than the *Eco*RI site. (**B**) Asterisks indicate point mutations, from CC to TT at positions –50 and –51, introduced into the L/S3 construct.

Several constructs disrupted the consensus SP1 site and a poly(dC) region that is similar to the consensus SP1 binding sequence, but showed normal levels of promoter activity (constructs L/S2, 19 and 3; Fig. 3A). To examine whether these two potential SP1 binding sites could be redundant in this context, a construct was created in which both of these sequences were mutated; indeed, the double mutant construct showed a significant decrease in promoter activity (Fig. 3B). This result suggests either that the double mutation results in a conformational change in the minimal promoter which is inhibitory to transcription or that some necessary transcription factor is sensitive to changes at the two regions together but not to changes at either position separately.

Electrophoretic mobility shift assays. In order to identify proteins involved in *XIST* minimal promoter activity, various sequences within the minimal promoter were subjected to DNA band shift analysis using HeLa cell nuclear extracts. Double-stranded DNA probes containing sequences identified by the mutagenesis screen as being important for transcription were subjected to electrophoretic mobility shift analysis in order to identify the proteins that bind to these regions. Probe 1 corresponds to the L/S7–L/S6 region, probe 2 to the L/S12 region and probe 3 to the L/S17 region (see Materials and Methods).

As seen in Figure 4A, *XIST* probe 1 displayed two band shifts with HeLa nuclear extract which were both competed away by excess unlabeled probe 1, but not by probe 2. Because the probe 1 sequence is similar to the consensus SP1 binding sequence, we also tested the ability of an unlabeled SP1 consensus oligo to compete with the probe 1 shifts. Indeed, as seen in Figure 4A, both shifts could be effectively competed by the SP1 consensus oligo, but not by a mutant SP1 oligo; this result suggests that *XIST* probe 1 is binding SP1. This conclusion was verified using anti-SP1 antibodies. As shown in Figure 4B, the shift seen for probe 1 with HeLa nuclear extract is similar to that seen with purified, recombinant SP1 protein. The addition of anti-SP1 antibody to the binding reaction results in a supershift in both the HeLa extract and the purified SP1 reactions, while the addition of a control antibody (anti-USF; see Materials and Methods) had no such effect. Notably, the results obtained with probe 1 were identical to those obtained with the SP1 consensus oligonucleotide. These results confirm that the region of the *XIST* minimal promoter represented by probe 1 binds SP1. However, since the HeLa extract shifts are complex and since there is residual material that is not supershifted by anti-SP1 (Fig. 4B), it is possible that additional nuclear proteins also bind this region of the *XIST* promoter.

The second region of the *XIST* promoter tested for protein binding was identified as being critical for promoter activity by construct L/S12. The normal sequence resembles the consensus binding site for the TATA binding protein (TBP), TATAAA, and is located in an appropriate position to be a functional TATA box (37,38). This led us to investigate the ability of this sequence to bind to TBP. As shown in Figure 4C, *XIST* probe 2 produces a complex band shift when incubated with HeLa nuclear extract. This shift can be competed away by excess unlabeled probe 2 or by excess unlabeled TBP consensus oligonucleotide much more efficiently than by excess unlabeled probe 1. This indicates that probe 2 binds TBP or a related protein. The TBP consensus oligonucleotide itself produces a much more intense shift than does *XIST* probe 2 (Fig. 4C), indicating that TBP binds to its consensus oligonucleotide with a much higher affinity than it does to probe 2. Probe 1 is also able to compete away the shift,

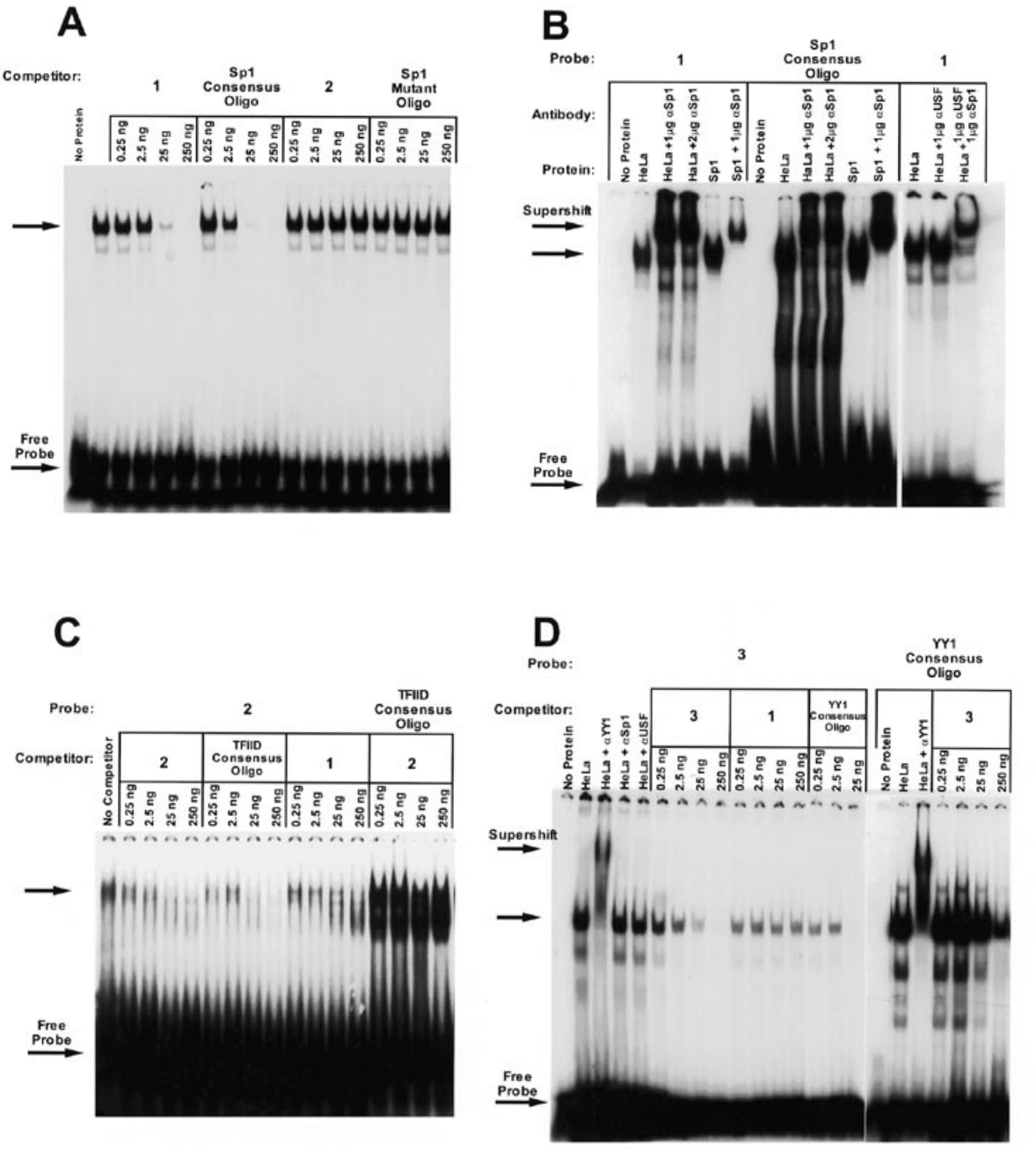

Figure 4. Electrophoretic mobility shift assays. Approximately 100 pg double-stranded DNA probe were radioactively labeled, incubated with 5µg HeLa nuclear extract or indicated recombinant protein and electrophoresed through a non-denaturing acrylamide gel. DNA–protein interactions are seen as a retardation of the mobility of the radioactive probe through the gel and are indicated with an arrow. Unlabeled competitor oligonucleotides were added to binding reactions in the amounts indicated. Probe 1 contains the sequence mutated in mutagenesis constructs L/S7 and L/S6, extending from positions –84 to –60. Probe 2 contains sequences mutated in construct L/S12 and extend from positions –37 to –16. Probe 3 contains sequences mutated in construct L/S17, extending from position –9 to +12. (**A**) Probe 1 forms a complex when incubated with HeLa nuclear extract. This complex is competed away by excess unlabeled probe 1 and SP1 consensus oligonucleotide, but not by probe 2 or a mutated form of the SP1 oligonucleotide that no longer binds SP1. (**B**) Probe 1 is supershifted by an antibody against SP1, but not by an antibody against the transcription factor USF. The supershift pattern is identical to that seen for the SP1 consensus probe. (**C**) The mobility shift of probe 2 incubated with 0.5 µg HeLa nuclear extract is competed away by excess unlabeled probe 2 and excess TFIID consensus oligonucleotide, but not as well by excess probe 1. At right, a band shift produced by the TFIID consensus oligonucleotide is considerably darker than that produced by probe 2 and is not competed away by excess unlabeled probe 2. (**D**) The mobility shift of probe 3 incubated with 0.5 µg HeLa nuclear extract is competed by excess unlabeled probe 3 or YY1 consensus oligonucleotide, but not by excess probe 1 or 2. A supershift is produced when an anti-YY1 antibody is included in the binding reaction. No supershift is produced by anti-SP1 or anti-USF antibodies. At right the mobility shift produced with the YY1 consensus oligonucleotide is competed by excess unlabeled probe 3 and shows a supershift very similar to that seen for probe 3.

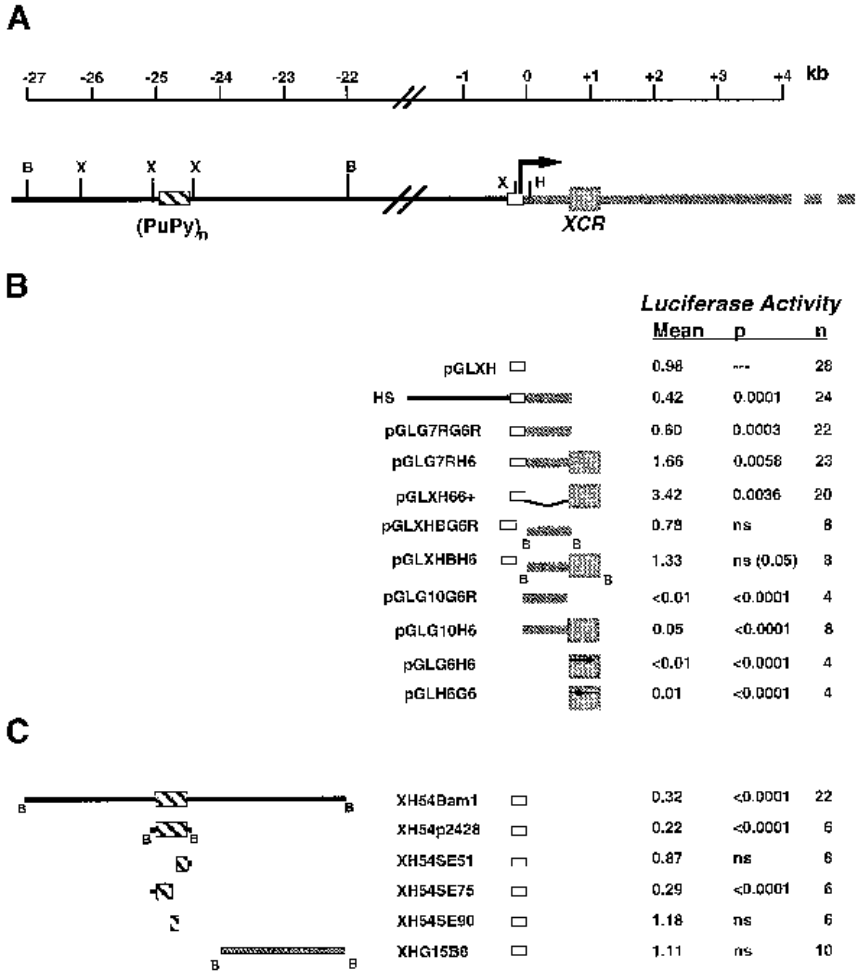

Figure 5. Effect of two repeat sequences on minimal promoter activity. Constructs were transiently transfected into 293 cells. Statistics are as in Figure 3. All luciferase values are normalized to the average of the pGLXH values for each experiment. (**A**) A schematic of the 5′-end of the *XIST* gene, with the minimal promoter shown as an open box, the XCR shown as a grey box and the PuPy repeat shown as a hatched box. The transcription initiation site is indicated by a large arrow. Selected restriction sites are as in Figure 2. (**B**) Effect on reporter gene activity of cloning the XCR into plasmid pGLXH, which contains the minimal promoter (open box) upstream of the promoterless luciferase gene. Potential regulatory sequences shown flanked by the letter B are cloned into a distal site on the plasmid that is not within the transcript. For pGLXHBH6, the *P* value in parentheses reflects comparison with pGLXHBG6R and demonstrates modest activity of the XCR sequences in this context. The nucleotide positions contained within each promoter construct are: HS, –923 to +324; pGLG7RG6R, –211 to +324; pGLG7RH6, –211 to +767; pGLXH66+, –211 to +31 and +304 to +767; pGLG10G6R, +29 to +324; pGLG10H6, +29 to +767; pGLG6H6 and pGLG6R6, +304 to +767 in both orientations. (**C**) Promoter constructs containing upstream sequences cloned into a distal site on the pGLXH plasmid containing the minimal promoter (open box). The hatched box represents PuPy sequences and the gray line represents a control *Bam*HI fragment located ~15 kb upstream of *XIST*.

but only at high concentrations, perhaps reflecting the known non-specific DNA binding ability of TBP (39).

The third sequence identified in the saturation mutagenesis experiment is contained within probe 3 and is similar to the consensus binding sequence of the transcriptional activator YY1. As shown in Figure 4D, a shift is obtained when *XIST* probe 3 is incubated with HeLa nuclear extract and this shift is competed away by excess unlabeled probe 3 or YY1 consensus oligonucleotide, but not by excess unlabeled probe 1 or probe 2. That this sequence is binding the initiator protein YY1 was confirmed using anti-YY1 antibodies (Fig. 4D). Thus, the decrease in luciferase activity seen in the L/S17 mutant construct (Fig. 3A) indicates that YY1 acts as a transcriptional activator in this context.

Two different repeat elements modify *XIST* minimal promoter activity *in vitro*

As determined above, expression from the *XIST* minimal promoter is driven by binding of common transcription factors in several different cell lines. Sequences outside this minimal promoter, therefore, may give the *XIST* gene its unique transcription pattern. In an attempt to identify such sequences, we cloned a series of restriction fragments from a λ phage contig including the human *XIST* gene and extending >50 kb upstream into a reporter plasmid in which the luciferase gene is driven by the *XIST* promoter (construct pGLXH, Fig. 2). Numerous fragments tested failed to produce a significant effect on minimal promoter activity when compared with the pGLXH construct, including

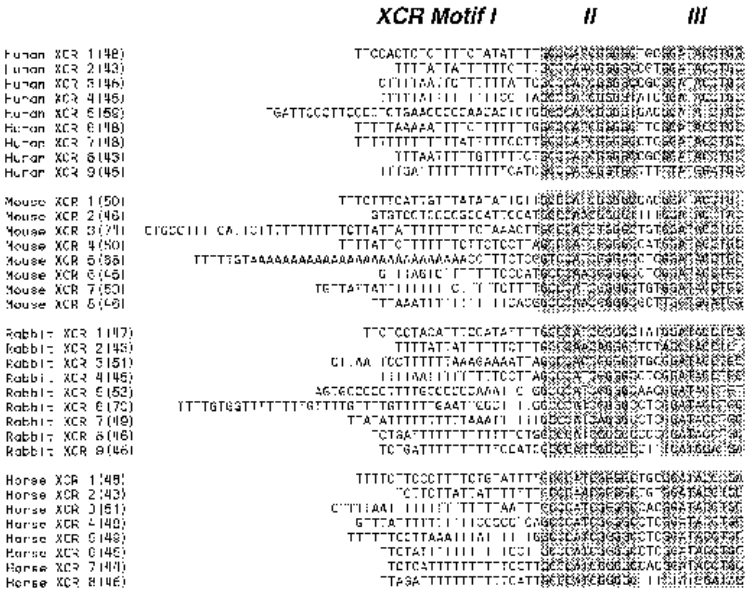

Figure 6. *XIST* conserved repeats (XCR). Human, mouse, rabbit and horse 5′ tandem repeats are aligned and begin at positions +309, +292, +303 and +298 of the human, murine, lepine and equine *XIST*/*Xist* genes respectively. XCR sequences are listed 5′→3′ with no gaps between repeats. The length of the individual repeat monomers is indicated within parentheses after the repeat number. Boxes indicate two highly conserved 12 and 10 bp cores (motifs II and III) within the overall repeat.

one containing the first intron. However, two sequences were found which did alter promoter activity significantly (Fig. 5A).

The 5′-end of the transcribed portion of *XIST* contains a series of nine tandem copies of a repeated motif that comprises a promising candidate region for a functional domain within the *XIST* transcript because of their high degree of conservation among eutherians (15,27). The repeats, designated XCR 1–9 (for *XIST* conserved repeats), consist of a highly conserved GC-rich core sequence (motifs II and III, separated by a more variable 3 bp 'hinge'), punctuated by a variable length spacer region (motif I) that is highly T rich (Fig. 6). The XCR sequences are located ~300 bp downstream of the transcription start site for all four homologs sequenced and are arranged in a head-to-tail fashion. In human *XIST* the overall structure is $(I–II–III)_9$, although motif III in XCR 9 is divergent. Motif I varies in length from 18 to 34 bp and in AT-richness from 74 to 94% (except for one copy that is 50% AT), with significant overall strand asymmetry, the transcribed strand being on average 84% pyrimidines. The same general structure is conserved in the mouse, rabbit and horse *Xist* genes, although the mouse and horse genes have only eight XCR copies. The consensus sequences of motifs II and III are identical in the four species (GCCCATCGGGGC and GGATACCTGC), with only ~4% divergence from the consensus among the 34 copies of motif II and the 30 copies of motif III. In each species, the 3′-most copy of motif III is divergent from the consensus, but is nonetheless conserved (75% identity) among the four genes (Fig. 6). Such high levels of primary sequence conservation are suggestive of functional significance.

Because the XCR sequences are present within the 5′-end of the *XIST* transcript, we wanted to determine whether the repeats would have any effect on transcript levels when present in the 5′-end of a reporter transcript. When tested in the 5′-untranslated portion of the luciferase reporter gene, the XCR sequences resulted in a >3-fold increase in luciferase activity ($P < 0.004$; Fig. 5B). This effect was seen in the female embryonic kidney cell line (293) as well as in mouse fibroblasts (data not shown). A similar but somewhat smaller effect was observed when a larger fragment containing transcribed sequences from the 5′-end of *XIST* between the minimal promoter and the repeats was included in the reporter construct ($P < 0.006$; Fig. 5B). As only a modest effect was observed when the repeats were cloned into the distant *Bam*HI site of the pGLXH construct, the data indicate that the effect of the XCR sequences is strongest when part of the transcript itself, consistent with a possible post-transcriptional effect. The possibility that the repeats contain promoter activity of their own was investigated by cloning the repeats alone, in either orientation, in front of the promoterless luciferase gene; however, no luciferase activity was observed in cells transfected with these constructs (Fig. 5B). The XCR sequences thus provide a moderate but significant enhancing activity, the effect of which is clearly strongest when present in the 5′-untranslated region.

A second region which influences minimal promoter activity *in vitro* was found within a 5.6 kb *Bam*HI fragment located ~25 kb upstream of the *XIST* gene. This sequence caused a significant reduction in luciferase activity in transient transfections of both male and female cell lines when present on the pGLXH plasmid, while other upstream genomic fragments had no such effect (Fig. 5C). Upon determination of the sequence of this *Bam*HI fragment, it was found that the fragment contains a stretch of alternating purine–pyrimidine (PuPy) repeats. This sequence extends for ~450 bp with only 14 single base interruptions in the strict purine–pyrimidine repeat structure. The repeat region was PCR amplified and cloned into the *Bam*HI site of the pGLXH promoter construct. A similar level of reduction in promoter activity was found with just the PuPy repeat as was found with the entire 5.6 kb *Bam*HI fragment, indicating that the repeats are responsible for the observed transcriptional repression. Two *Sna*BI restriction sites lie within the repeat region and were used to generate constructs containing different portions of the repeat array directly upstream of the minimal promoter. A construct

with 299 bp of repeat greatly inhibited luciferase activity, whereas constructs with 166 or 25 bp of PuPy repeat showed normal luciferase activity (Fig. 5C), illustrating that >166 bp of the *XIST* PuPy repeat is required to promote silencing when located directly next to a promoter in this *in vitro* assay.

DISCUSSION

Expression of the *XIST* gene and accumulation of *XIST* RNA only occurs on the inactive X chromosome or on chromosomes that are programmed to undergo X chromosome inactivation (14,15,17,18,40–42). Thus factors controlling *XIST* expression are directly involved in, or are influenced by, the X inactivation process. By understanding *XIST* transcriptional regulation, we hope to gain insights into the nature of the initiation of X inactivation. Here we report the identification and characterization of the human *XIST* minimal promoter and the identification of two repeat elements that modulate minimal promoter activity in an *in vitro* assay.

The *XIST* promoter is constitutively active and binds common transcription factors

XIST RNA is detected in significant amounts only from inactive X chromosomes and not from active X chromosomes. *A priori*, such a pattern of differential gene expression could reflect one of two alternative possibilities (Fig. 7). First, the *XIST* promoter could be a conditional one and require transcriptional activation by factors specific to X chromosomes chosen to become an inactive X. Alternatively, *XIST* could be constitutively active on all X chromosomes prior to inactivation (as suggested by the studies of Panning and Jaenisch; 43) and require transcriptional repression on the single X chromosome in males and the active X in females. Under such a model, the high levels of *XIST* RNA found associated specifically with inactive X chromosomes (15,16) may reflect up-regulation of the *XIST* promoter from a basal, constitutive level (the 'pre-inactivation state'; 44) and/or stabilization of *XIST* transcripts by factors involved in assembly of the inactive X Barr body complex. The fact that we readily detect promoter activity in a number of different cell lines as well as in transgenic mouse lines indicates that the *XIST* minimal promoter is constitutively capable of supporting transcription and thus must be silenced on the active X chromosome. This observation is consistent with models in which a developmental factor(s) acts to 'mark' or 'block' *XIC* on a single active X, regardless of the total number of X chromosomes present (19,44). Indeed, studies of *XIST*/*Xist* methylation in the region of the minimal promoter (28,45–47) have implicated DNA methylation in the silencing of *XIST* on the active X chromosome, perhaps as part of the hypothesized 'blocking' step.

Transcription factors responsible for *XIST* expression were identified through the use of saturation site-directed mutagenesis and gel mobility shift assays. An SP1 binding sequence centered around position –72 and the potential SP1 binding sequences at positions –50 to –43 (Fig. 1) show positioning expected for upstream control elements and were implicated by the mutagenesis studies (Fig. 3). The second identified sequence is very similar to the consensus binding sequence of the initiator protein YY1 and is located at the expected position for YY1 to bind as an activator of transcription (Fig. 1; 37); this sequence was indeed found to bind the YY1 protein *in vitro* (Fig. 4D). The third sequence is located from position –26 to –31 with respect to the transcription start site, the location expected for a TATA box. Indeed, the sequence which overlaps this region, TTAAAG, is very similar to the TBP consensus binding sequence and was found to bind either TBP or some TBP-like protein in electric mobility shift assays (Fig. 4C). A recent characterization of the murine *Xist* minimal promoter identified the TTAAAG sequence as also being important for promoter activity (48).

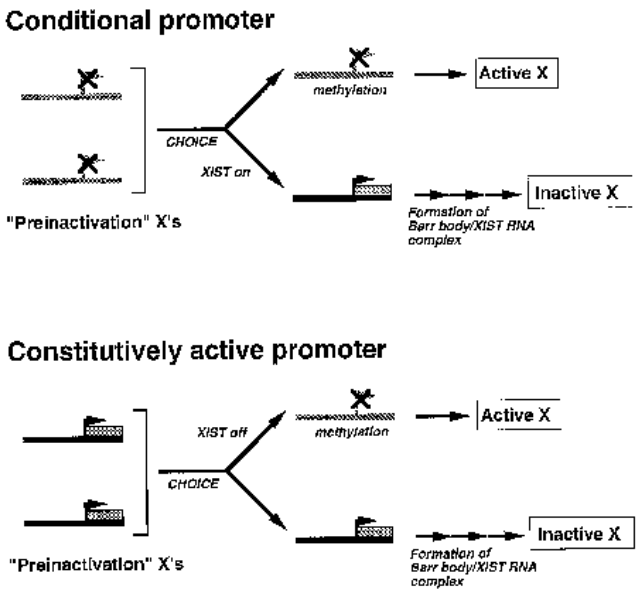

Figure 7. Two models of events at the *XIST* promoter in early development. Under the conditional promoter model (top) a key element in the choice between Xs in female cells is activating the *XIST* gene on the X chromosome chosen to become the inactive X. Under the constitutively active model (bottom) both copies of *XIST* are transcribed prior to X inactivation and the key step in the choice between the two Xs is repression of *XIST* on the X that will be the active X. In both models additional unknown steps and/or factors are required for accumulation of *XIST* RNA on the inactive X chromosome (15,16) and formation of the Barr body complex.

The transient transfection data and the results obtained in the analysis of protein interactions with the minimal promoter support the constitutively active model for the *XIST* promoter shown in Figure 7. The minimal promoter directs high levels of transcription in all cell types tested using common transcription factors. Similarly, luciferase activity is readily detectable in transgenic mouse lines in which the human *XIST* promoter is driving luciferase expression. Thus, both the *in vitro* and *in vivo* data provide no evidence for a DNA binding factor responsible for the inactive X-specific expression characteristic of the *XIST* gene, as required by the conditional *XIST* promoter model.

These results are interesting in the light of three instances where *Xist*-containing sequences have been used to create transgenic mice (20–24). Despite the fact that the *XIST*/*Xist* minimal promoters are constitutively active (Fig. 1), transgenic lines were obtained in two of these reports in which transgenic *Xist* expression was not detected, though the input YACs were apparently intact (20,21). The variable data on such transgenes may implicate *cis*-acting chromatin elements (perhaps as part of a *XIC*/*Xic* recognition element) that are not always maintained in a proper context in the transgenic lines. Both Jaenisch and colleagues (22,23) and Ashworth and colleagues (24) have

detected *Xist* expression and accompanying spread of *Xist* RNA in transgenic ES lines. However, the multicopy nature of those transgenes complicates conclusions regarding the possible existence or location of such *cis*-acting elements within *XIC*/*Xic*.

Minimal promoter activity is modulated by two repeated DNA elements

We have identified two sequences capable of influencing the activity of the *XIST* minimal promoter in reporter gene constructs, both of which consist of repeated DNA (Fig. 5). The XCR sequences present ~300 bp downstream from the transcription start site act to stimulate minimal promoter activity, perhaps by a post-transcriptional mechanism. While we have been unable thus far to demonstrate the existence of protein binding to the XCR using HeLa cell nuclear extracts (unpublished data), it is conceivable that a developmentally regulated factor, present, for example, at the time of initiation of X inactivation, may bind the XCR sequences and thereby enhance *XIST* expression and/or stabilize *XIST* RNA *in vivo*.

The upstream PuPy repeat sequences are also implicated in control of *XIST* expression and cause a 70–80% reduction of promoter activity in transient transfection assays. Similar PuPy sequences at other loci have been shown to form a right handed Z-DNA structure under physiological conditions *in vitro* (49–52). The mode of action of the PuPy sequence in inhibition of *XIST* promoter activity is unclear. While the sequence of the repeat is not unique, the *XIST* PuPy repeat is nearly twice the size of other known PuPy repeats (52). Should this sequence form a Z-DNA structure, it would represent a large segment of non-B-DNA and could conceivably direct formation of a repressive chromatin structure in the vicinity of the *XIST* gene and *XIC*. Further work will be necessary to determine whether either of these repeat sequences influence *XIST* expression levels *in vivo* and are therefore true *cis*-acting elements involved in the X inactivation process.

Initiation events at the X inactivation center

All X chromosomes in excess of one present in a diploid cell are inactivated. This suggests that a cell is able to recognize all *XIC* loci (i.e. X chromosomes) within that cell, 'choose' one *XIC* and 'block' the chosen *XIC* to keep that chromosome active and ensure that all remaining *XIC* loci inactivate the chromosomes on which they lie.

Penny *et al*. (18) have recently described experiments in which the proximal minimal promoter and entire first exon of one *Xist* allele were deleted in female ES cells. When cells containing this deleted *Xist* gene were induced to differentiate, the chromosome harboring the deleted allele could be recognized and 'chosen', but could not become inactive. This demonstrates that the *Xist* gene is absolutely required for X inactivation to occur and implies that sequences involved in *Xic* recognition and choice were not affected by the deletion (19). Panning and Jaenisch (43) have observed low levels of biallelic *Xist* expression in undifferentiated ES cells which, upon differentiation, switched to monoallelic expression. This is consistent with the initiation step of X inactivation consisting of silencing one *Xist* allele and up-regulating the other and supports the constitutive promoter model (Fig. 7). Together with the data reported here, these two reports support the idea that there may be at least three critical and possibly distinct types of X-linked sequences involved in X chromosome inactivation; those involved in recognition and choice of one X chromosome, those involved in blocking *XIST*/*Xist* transcription from that allele on the X chosen to be the active X and those responsible for the control and/or stabilization of *XIST*/*Xist* expression on the other X (or Xs) that is/are subsequently inactivated. While the experiments reported here and elsewhere (18) clarify those sequences responsible for *XIST*/*Xist* transcription, they suggest that the putative *XIC* recognition elements lie, at least in part, outside the examined regions.

While the nature of the 'blocking' step is not known, a likely candidate mechanism is differential DNA methylation. The importance of DNA methylation in silencing the *Xist* gene is demonstrated by the fact that mice lacking maintenance DNA methyltransferase activity show inappropriate *Xist* expression in somatic cells (43,53). Further, the fact that the minimal promoter and XCR sequences are differentially methylated on the active and inactive X chromosomes (28,45–47) indicates that although this region is not important in choosing, it may be the region on which the 'block' (in the form of differential DNA methylation) is imposed, thus preventing *XIST*/*Xist* expression and allowing the chromosome to remain active. The identification of a naturally occurring point mutation in the *XIST* promoter that segregates with non-random X chromosome inactivation in a human family (54) further supports the possibility that the minimal promoter region is important in the very early stages of X chromosome inactivation and may be involved in choosing between the two X chromosomes in female cells and/or establishing the blocking signal.

Epigenetics and *XIST* transcription

How might DNA methylation, the minimal promoter and *cis*-acting elements interact to achieve the unique expression pattern observed for *XIST*? SP1 has been shown to bind to its consensus site and to promote transcription irrespective of the methylation status of the promoter sequence (55–57). Thus, silencing of *XIST* on the active X chromosome cannot be achieved through exclusion of SP1 from promoter sequences by methylation alone. Further, DNA gel mobility assays using methylated and unmethylated minimal promoter probes showed no differences in mobility shift patterns (data not shown), providing further evidence that DNA methylation itself does not affect binding of proteins to the *XIST* minimal promoter and, therefore, is not likely to be directly responsible for silencing of *XIST* on the active X chromosome in somatic cells. The XCR region is known to be differentially methylated in somatic cells (28; Fig. 7) and has a number of conserved CpG dinucleotides within the motif II and III consensus sequences (Fig. 6). Whether such methylated sequences might inhibit *XIST* transcription is unclear at present.

That expression of the *XIST*/*Xist* genes should be under the control of at least one distant, *cis*-acting DNA element is expected from studies of the *Xce* locus in mice (7). Alleles at *Xic*-linked *Xce* influence the randomness of X chromosome inactivation in mice as well as influencing the steady-state levels of *Xist* RNA in somatic cells. While *Xce* has been found to be genetically distinct from the *Xist* gene in one mouse strain (58), it remains possible that an interaction between the *Xist* promoter and *cis*-acting element(s) act to achieve the *Xce* effects. Indeed, we have previously identified strain-specific sequence variations in the murine XCR region (28) which could contribute to this effect.

Recent work has also identified *Xce* allele-specific methylation differences in a region lying distal to the murine *Xist* gene, providing a candidate region for the *Xce* locus (59). It will be interesting to learn whether this region contains *cis*-acting elements that, like the PuPy repeats examined here, influence *Xist* minimal promoter activity. Further work on the interaction between the minimal promoter and *cis*-acting elements will be necessary to determine what role these sequences play in the processes of *XIST* transcription, *Xce* effects and, subsequently, X chromosome inactivation.

ACKNOWLEDGEMENTS

We thank T.Mourton for technical advice and for creating the transgenic mice at the Case Western Reserve University Transgenic Mouse Facility. We also thank Drs T.Magnuson, R.Myers and D.Setzer for useful discussions, Mary Schuler for technical advice and Dr A.Bird for useful comments on the manuscript. This work was supported in part by a research grant from the National Institutes of Health to H.F.W. (GM45441). R.M.P. was supported by a research fellowship from the Howard Hughes Medical Institute.

REFERENCES

1 Cattanach,B.M. (1970) *Genet. Res*., **16**, 293–301.
2 Cohen,M.M., Lin,C.C., Sybert,V. and Orecchio,E.J. (1972) *Am. J. Hum. Genet*., **24**, 583–597.
3 Lyon,M.F. (1963) *Proceedings of the International Conference on Congenital Malformations.* 2nd International Medical Congress Ltd, New York, NY, pp. 67–68.
4 Russell,L.B. (1964) *Trans. NY Acad. Sci. Ser. II*, **26**, 726–736.
5 Therman,E., Sarto,G.E. and Patau,K. (1974) *Chromosoma*, **44**, 361–366.
6 Brown,C.J., Lafreniere,R.G., Powers,V.E., Sebastio,G., Ballabio,A., Pettigrew,A.L., Ledbetter,D.H., Levy,E., Craig,I.W. and Willard,H.F. (1991) *Nature*, **349**, 82–84.
7 Heard,E. and Avner,P. (1994) *Hum. Mol. Genet*., **3**, 1481–1485.
8 Leppig,K.A., Brown,C.J., Bressler,S.L., Gustashaw,K., Pagon,R.A., Willard,H.F. and Disteche,C.M. (1993) *Hum. Mol. Genet*., **2**, 883–888.
9 Cattanach,B.M., Pollard,C.E. and Perez,J.N. (1969) *Genet. Res*., **14**, 223–235.
10 Cattanach,B.M. (1975) *Annu. Rev. Genet*., **9**, 1–18.
11 Cattanach,B.M. and Williams,C.E. (1972) *Genet. Res*., **19**, 229–240.
12 Cattanach,B.M. and Papworth,D. (1981) *Genet. Res*., **38**, 57–70.
13 Cattanach,B.M., Rasberry,C. and Andrews,S.J. (1989) *Mouse Newsl*., **83**, 165.
14 Brown,C.J., Ballabio,A., Rupert,J.L., Lafreniere,R.G., Grompe,M., Tonlorenzi,R. and Willard,H.F. (1991) *Nature*, **349**, 38–44.
15 Brown,C.J., Hendrich,B.D., Rupert,J.L., Lafreniere,R.G., Xing,Y., Lawrence,J. and Willard,H.F. (1992) *Cell*, **71**, 527–542.
16 Clemson,C.M., McNeil,J.A., Willard,H.F. and Lawrence,J.B. (1996) *J. Cell. Biol*., **132**, 259–275.
17 Kay,G.F., Penny,G.D., Patel,D., Ashworth,A., Brockdorff,N. and Rastan,S. (1993) *Cell*, **72**, 171–182.
18 Penny,G.D., Kay,G.F., Sheardown,S.A., Rastan,S. and Brockdorff,N. (1996) *Nature*, **379**, 131–137.
19 Lyon,M.F. (1996) *Nature*, **379**, 116–117.
20 Heard,E., Kress,C., Mongelard,F., Courtier,B., Rougeulle,C., Ashworth,A., Vourc'h,C., Babinet,C. and Avner,P. (1996) *Hum. Mol. Genet*., **5**, 441–450.
21 Matsuura,S., Episkopou,V., Hamvas,R. and Brown,S.D.M. (1996) *Hum. Mol. Genet*., **5**, 451–459.
22 Lee,J.T., Strauss,W.M., Dausman,J.A. and Jaenisch,R. (1996) *Cell*, **86**, 83–94.
23 Lee,J.T. and Jaenisch,R., (1997) *Nature*, **386**, 275–279.
24 Herzing,L., Romer,J.T., Horn,J.M. and Ashworth,A. (1997) *Nature*, **286**, 272–275.
25 Eicher,E.M. (1970) *Adv. Genet*., **15**, 175–257.
26 Brockdorff,N., Ashworth,A., Kay,G.F., Cooper,P., Smith,S., McCabe,V.M., Norris,D.P., Penny,G.D., Patel,D. and Rastan,S. (1991) *Nature*, **351**, 329–331.
27 Buzin,C.H., Mann,J.R. and Singer-Sam,J. (1994) *Development*, **120**, 3529–3536.
28 Hendrich,B., Brown,C. and Willard,H.F. (1993) *Hum. Mol. Genet*., **2**, 663–672.
29 Uberbacher,E.C. and Mural,R.J. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 11262–11265.
30 McMahon,A.P. and Bradley,A. (1990) *Cell*, **62**, 1073–1085.
31 Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) *J. Mol. Biol*., **215**, 403–410.
32 Willard,H.F., Brown,C.J., Carrel,L., Hendrich,B. and Miller,A.P. (1993) *Cold Spring Harbor Symp. Quant. Biol*., **58**, 315–322.
33 Kadonaga,J.T., Jones,K.A. and Tjian,R. (1986) *Trends Biochem. Sci*., **11**, 20–23.
34 Becker,K.G., Jedlicka,P., Templeton,N.S., Liotta,L. and Ozato,K. (1994) *Gene*, **150**, 259–266.
35 Shi,Y., Seto,E., Chang,L.-S. and Shenk,T. (1991) *Cell*, **67**, 377–388.
36 Seto,E., Shi,Y. and Shenk,T. (1991) *Nature*, **354**, 241–245.
37 Hahn,S., Buratowshi,S., Sharp,P. and Guarente,L. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5718–5722.
38 Hahn,S. (1992) *Curr. Biol*., **2**, 152–154.
39 Conaway,R.C. and Conaway,J.W. (1993) *Annu. Rev. Biochem*., **62**, 161–190.
40 McCarrey,J.R. and Dilworth,D.D. (1992) *Nature Genet*., **2**, 200–203.
41 Salido,E.C., Yen,P.H., Mohandas,T.K. and Shapiro,L.J. (1992) *Nature Genet*., **2**, 196–199.
42 Brockdorff,N., Ashworth,A., Kay,G.F., McCabe,V.M., Norris,D.P., Cooper,P.J., Swift,S. and Rastan,S. (1992) *Cell*, **71**, 515–526.
43 Panning,B. and Jaenisch,R. (1996) *Genes Dev*., **10**, 1991–2002.
44 Willard,H.F. (1996) *Cell*, **86**, 5–7.
45 Norris,D.P., Patel,D., Kay,G.F., Penny,G.D., Brockdorff,N., Sheardown,S.A. and Rastan,S. (1994) *Cell*, **77**, 41–51.
46 Ariel,M., Robinson,E., McCarrey,J.R. and Cedar,H. (1995) *Nature Genet*., **9**, 312–315.
47 Zuccotti,M. and Monk,M. (1995) *Nature Genet*., **9**, 316–320.
48 Pillet,N., Bonny,C. and Schorderet,D.F. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 12515–12519.
49 Collier,D.A., Griffin,J.A. and Wells,R.D. (1988) *J. Biol. Chem*., **263**, 7397–7405.
50 Naylor,L.H. and Clark,E.M. (1990) *Nucleic Acids Res*., **18**, 1595–1601.
51 Thomas,M.J., Freeland,T.M. and Strobl,J.S. (1990) *Mol. Cell. Biol*., **10**, 5378–5387.
52 Yu,M. and Edenberg,H.J. (1995) *Arch. Biochem. Biophys*., **316**, 407–412.
53 Beard,C., Li,E. and Jaenisch,R. (1995) *Genes Dev*., **9**, 2325–2334.
54 Plenge,R.M., Hendrich,B.D. and Willard,H.F. (1995) *Am. J. Hum. Genet*., **57** (suppl.), A30.
55 Ben-Hattar,J., Beard,P. and Jiricny,J. (1989) *Nucleic Acids Res*., **17**, 10179–10190.
56 Harrington,M.A., Jones,P.A., Imagawa,M. and Karin,M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2066–2070.
57 Höller,M., Westin,G., Jiricny,J. and Schaffner,W. (1988) *Genes Dev*., **2**, 1127–1135.
58 Simmler,M.-C., Cattanach,B., Rasberry,C., Rougeulle,C. and Avner,P. (1993) *Mamm. Genome*, **4**, 523–530.
59 Courtier,B., Heard,E. and Avner,P. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 3531–3535.